Users Guide to Designing N-of-1 Trials

Chapter 1

Introduction to N-of-1 Trials: Indications and Barriers

Draft for Comment

The goal of evidence-based medicine (EBM) is to integrate research evidence, clinical judgment, and patient preferences in a way that maximizes benefits and minimizes harms to the individual patient. The foundational, gold standard research design in EBM is the randomized, parallel group clinical trial. However, the majority of patients may be ineligible for or unable to access such trials.[1] In addition, these clinical experiments generate average treatment effects, which may not apply to the individual patient. To generate individual treatment effects (ITEs), clinical investigators have taken several tacks, including subgroup analysis, matched pairs designs, and n-of-1 trials. Of these, n-of-1 trials provide the most direct route to estimating the effect of a treatment on the individual. In this chapter, we introduce n-of-1 trials by providing definitions and a rationale, delineating indications for use, describing key design elements, and addressing major opportunities and challenges.

Defining n-of-1 Trials

N-of-1 trials in clinical medicine are multiple crossover trials, usually randomized and often blinded, conducted in a single patient. As such, n-of-1 trials are part of a family of Single Case Designs that have been widely used in psychology, education, and social work. In the schema of Perdices et al., the Single Case Designs family includes case descriptions, non-randomized designs, and randomized designs.[2] N-of-1 trials are a specific form of randomized or balanced designs characterized by periodic switching from active treatment to placebo or between active treatments ("withdrawal-reversal designs"). N-of-1 trials were introduced to clinicians by Hogben and Sim as early as 1953,[3] but it took 30 years for the movement to find an effective evangelist.[4-6] Many of the pioneers of the movement established active n-of-1 trial units in academic centers, only to abandon them once funding was exhausted.[7] However, several units are still thriving, and over the past three decades over 2000 patients have participated in published n-of-1 trials; less than 10% chose treatments inconsistent with the results.

In contrast to parallel group trials, n-of-1 trials use *crossover* between treatments to address the problem of "patient by treatment interaction." This situation arises when characteristics of the individual affect whether Treatment A or Treatment B (which could be an active treatment, a placebo, or no treatment) delivers superior results. Also, by prescribing *multiple episodes* of treatment, n-of-1 trials increase precision of measurement and control for "treatment by time" interaction, the possibility that the relative effects of two treatments may vary over time.

Rationale for n-of-1 Trials in the Era of Patient-Centered Care

The success of an n-of-1 trial largely depends on the collaboration and commitment of both clinician and patient. *Clinicians* must explain the process to their patients, collaborate with patients in developing outcome measures most appropriate to the individual, monitor patients at regular intervals throughout the trial period, and evaluate and explain what the results of the trial mean. *Patients* participating in n-of-1 trials must be involved in selecting therapies for evaluation, record outcomes, engage in understanding their own data, and share in treatment decision-making. As the centerpiece of patient-centered care, patient engagement has been shown to result in better health outcomes among patients with chronic illness.[8] Nikles et al reported that patients who had completed an N-of-1 trial had a greater understanding and awareness of their condition, and felt a greater sense of control when it came to decisions about their health.[9].

Beyond their potential for promoting patient-centered care, n-of-1 trials may have additional, pragmatic significance. With escalating drug costs, health care systems are struggling to provide cost-effective therapies for patients. N-of-1 trials offer an objective way of determining individual response to therapy; if two therapeutic options are shown to have equivalent effectiveness in a given individual, the least costly option could be chosen. This approach to comparative effectiveness could apply to different classes of medications, as well as formal

assessment of the bioequivalence of generic and proprietary pharmaceuticals. Considering that n-of-1 trials are particularly suited to chronic conditions, the savings to the health care system could be substantial.

Indications, Contraindications, and Limitations

N-of-1 trials are indicated whenever there is substantial uncertainty regarding the comparative effectiveness of treatments being considered for an individual patient. Uncertainty can result from a general lack of evidence, as when the relevant parallel group randomized controlled trials (RCTs) have not been conducted, when the existing evidence is in conflict, or when the evidence is not relevant to the patient at hand.[10] Uncertainty may also result from the presence of substantial heterogeneity of treatment effects (HTE) across patients that cannot be easily predicted from available prognostic factors. HTE is the variance of ITEs across patients, where the ITE is the difference in effects (net benefits) between Treatment A and Treatment B for an individual patient.[11] The extent of HTE for common conditions and treatments is not well characterized, but some analyses suggest it is substantial.[12-16]

N-of-1 trials are applicable to chronic, stable or slowly progressive conditions that are either symptomatic or for which a valid biomarker has been identified. Unless recurrent, acute conditions lend no opportunity for multiple crossovers. Rapidly progressive conditions (or those prone to sudden, catastrophic outcomes such as stroke or death) are not amenable to the deliberate experimentation of n-of-1 trials. Asymptomatic conditions make outcomes assessment difficult. An exception would be when a valid biomarker exists.[17] Examples of such biomarkers might include blood pressure or LDL cholesterol in heart disease, sedimentation rate in some chronic autoimmune diseases, or intraocular pressure in glaucoma.

For practical reasons, treatments to be assessed in n-of-1 trials should have relatively rapid onset and brisk washout (i.e. few lasting carry-over effects). Treatments with a very slow onset

of action (e.g. methotrexate in rheumatoid arthritis) could necessitate treatment period lengths that exceed the patience of the average patient and clinician.[18] On the other hand, treatments with prolonged carryover effects would require a substantial washout period to distinguish readily between the effect of the currently administered treatment and the effect of the last treatment.[5] Some patient groups (e.g., patients with rare diseases) may be particularly motivated to participate in n-of-1 trials owing to the paucity of other evidence needed to substantiate treatment effect.

Major Design Elements of N-of-1 Trials

The major design elements of n-of-1 trials are balanced sequence assignment, blinding, and systematic outcomes measurement. Before introducing these elements, we begin by describing standard clinical practice.

Standard clinical practice. In ordinary practice, the clinician prescribes treatment and asks that the patient return for follow-up. At the follow-up encounter, the clinician asks the patient if he or she is improving. If the patient responds positively, the treatment is continued. If not, the clinician and patient discuss alternative strategies such as a dose increase, switching to a different treatment, or augmenting with a second treatment. This process continues until both agree that a satisfactory outcome has been achieved, until intolerable side effects occur, or until no further progress seems possible. Although treatments are administered in sequence, there is no systematic repetition of prior treatments (replication) and the treatment assignment sequence is based on physician and patient discretion (not randomized or balanced). Neither clinician nor patient is blinded. Typically, there is no systematic assessment of outcomes. As a result, it is easy for both patient and clinician to be misled about the true effects of a particular therapy.

Take for example, Mr. J, who presents to Dr. Alveolus during the spring allergy season with a nagging dry cough that is worse at night. After ruling out drug effects and infection, Dr. Alveolus posits post-nasal drip as the cause of Mr. J's cough and prescribes diphenhydramine 25 mg each night. The patient returns in a week and notes that he's a little better, but the "cough is still there." Dr. Alveolus increases the diphenhydramine dose to 50mg, but the patient stops taking it after 3 days because of morning drowsiness. He returns complaining of the same symptoms 2 weeks later; the doctor prescribes cetirizine 10mg (a non-sedating antihistamine). Mr. J fills the prescription but doesn't return for follow-up until after the summer. "How did the second pill I prescribed work out for you," Dr. Alveolus asks. "I think it helped," Mr. J replies, "but after a while the cough just went away so I stopped taking it."

N-of-1 trial procedures in contrast to standard clinical practice. What if Mr. J and Dr. Alveolus were to acknowledge their uncertainty and elect to embark on an n-of-1 trial of diphenhydramine versus cetirizine for treatment of chronic cough presumed due to allergic rhinitis? They might agree:

- To administer diphenhydramine and cetirizine in a balanced sequence of 7 day treatment intervals* for a total of 8 treatment periods (4 periods on diphenhydramine, 4 periods on cetirizine, 56 days total), with no washout time in-between treatment periods;
- To ask the compounding pharmacist to place the medications in identical appearing capsules;
- To assess *benefits* using the average of Mr. J's rating of overall cough severity (1-5 scale) and Mrs. J's rating of nighttime cough severity (1-5 scale) and *harms* using a daytime sleepiness scale.

* A 7-day interval is chosen for convenience and because shorter intervals might introduce confounding by day-of-week effects (e.g., the difference between weekends and weekdays).

Their design (schematized in Figure 1) incorporates, respectively, balanced sequence assignment, blinding, and systematic outcomes assessment.

We now discuss these 3 elements in greater detail.

Balanced repeated sequences. In parallel group RCTs, randomization serves to maximize the likelihood of equivalence between treatment groups (in terms of both known and unknown prognostic factors). In n-of-1 trials, the aim is to achieve balance in the assignment of treatments over time so that treatment effect estimates are unbiased by time-dependent confounders. Randomization is one way of achieving such balance, but there are others. For example, the treatment sequence AAAABBBB offers no protection against a confounder whose effect on the outcome is linear with time (e.g., a secular trend). The paired design ABABABAB and the singly counterbalanced design ABBAABBA offer better protection against temporally linear confounders but are still vulnerable to non-linear confounding. It can be shown that the doubly counterbalanced design ABBABAAB defends against secular trends that are both linear and non-linear. Repetition is critical; by exposing an individual patient to the same treatment more than once, we can control (at least in part) for random fluctuations in the patient's experience over time. *Balanced assignment* (which may include randomization) minimizes bias in two ways. First, it helps to control for time-varying clinical and environmental factors that could affect the patient's outcome.[19,20] Some, but not all of these factors may be known to the patient and clinician in advance. For example, Mr. J might have decided to take diphenhydramine on weekends and cetirizine on weekdays. He might then be less prone to notice daytime sleepiness from diphenhydramine because he tends to sleep in on weekends. This would bias his assessment. Second, combined with blinding, randomization makes it more difficult to guess which treatment has been assigned.

[Place Figure 1 about here.]

The importance of a washout period separating active treatment periods in n-of-1 trials has been fiercely debated.[21] A washout period is theoretically important whenever lingering effects of the first treatment might influence outcome measurements obtained while on a subsequent treatment. Carry-over effects resulting from insufficient washout will often tend to reduce observed differences between treatments for placebo controlled trials. However, more complex interactions are possible. For example, if the benefits of a particular treatment wash out quickly but the risks of adverse treatment-related harm persist (think aspirin, which reduces pain over a matter of hours but increases risk of bleeding for up to 7 days), the likelihood of detecting net benefit will depend on the order in which the treatments are administered. Similar issues also apply to slow onset of the new treatment. A possible downside of a washout period is that the patient is forced to spend some time completely off treatment, which might be undesirable for patients who already receive some benefit from both treatments. For practical purposes, washout periods may not be necessary when treatment effects (e.g. therapeutic half lives) are short relative to the length of the treatment periods. Since treatment half-lives are often not well-characterized and vary among individuals, the safest course may be to choose treatment lengths long enough to accommodate patients with longer-than-average treatment half-lives and to take frequent (e.g., daily) outcome measurements. An alternative to the use of a "physical" washout is the use of "analytic washout", to address the effects of carryover and slow onset analytically. Further discussions are given in Chapter 4 (Statistics).

Some n-of-1 investigators have advocated for the use of run-in periods. In parallel group RCTs, a run-in period is a specified period of time after enrollment and prior to randomization that is allotted to further measure a participant's eligibility and commitment to a study.[22] In n-of-1 trials, a run-in period could also be used to differentiate "responders" from "non-responders" in an open-label (unblinded) situation or to initiate dose-finding.

Blinding. In parallel group RCTs, blinding of patients, clinicians, and outcomes assessors ("triple blinding") is considered good research practice. These trials aim to generate generalizable knowledge about the effects of treatment in a population. In drug and device trials, the prevailing consensus is that it is critical to separate out the biological activity of the intervention from non-specific (placebo) effects. (For a broader view, see Benedetti J Neurosci 2005).[23] In n-of-1 trials, the primary aim is usually different. Patients and clinicians participating in n-of-1 trials are likely interested in the net benefits of treatment *overall,* including both specific *and non-specific* effects. Therefore blinding may be less critical in this context. Nevertheless, expert opinion tends to favor blinding in n-of-1 trials whenever feasible.

However, just as in parallel group randomized trials, blinding is not always feasible. For example, in trials of behavioral interventions (e.g. bibliotherapy versus computer-based cognitive behavioral therapy for depression), patients will always know what treatment they are on. Furthermore, even for drug trials, few community practitioners have access to a compounding pharmacy that can safely and securely prepare medications to be compared in matching capsules. However, as in parallel group RCTs of behavioral interventions, the absence of blinding does not detract from the importance of the evaluation.

Systematic outcomes assessment. Evidence is accumulating that careful, systematic monitoring of clinical progress supports better treatment planning and leads to better outcomes. For example, home blood pressure monitoring results in better blood pressure control[24] and "treat-to-target" approaches based on PHQ-9 scores have worked well in depression.[25] In n-of-1 trials, systematic assessment of outcomes may well be the single most important design element. There are two issues to consider: 1) what data to collect and 2) how to collect it.

In designing an n-of-1 trial, participants (patients, clinicians, investigators) must first select outcome domains (e.g. specific symptoms, specific dimensions of health status, etc.) and then

specific measures tapping those domains. In so doing, they must balance a number of competing interests. For most chronic conditions, there are numerous potentially relevant outcomes. These may be condition-specific (e.g., low back pain intensity in chronic low back pain, diarrhea frequency in inflammatory bowel disease) or generic (e.g. health related quality of life). Clinicians, patients and service administrators may assign different priorities to different domains. For example, in chronic musculoskeletal pain, the patient may prioritize control of pain intensity or fatigue, the clinician may prioritize daily functioning, and Drug Enforcement Agency officials may prioritize minimizing opportunities for misuse of opiates. The primary purpose of most n-of-1 trials is to assist with individual treatment decisions. Therefore patient preferences are paramount. However, as prescribers of treatment, clinicians are essential partners, and their buy-in is essential. At times, both patients and clinicians may sometimes yield to other priorities, such as a desire to see their data combined with n-of-1 trial data from other, similar patients. Combining n-of-1 trials is performed for two purposes: 1) to take advantage of the experience of past similar patients to generate better treatment recommendations for the current n-of-1 participant, and 2) to produce generalizable results that might be applicable to future patients. Combining n-of-1 trial results using Bayesian techniques will be addressed in more detail in Chapter 4 (Statistics).

Once outcome domains have been identified, participants need to pick specific measures. When available, pre-existing measures known to possess high reliability and validity are preferable. However, sometimes an appropriate pre-existing measure cannot be found. In this case, n-of-1 participants must choose between measures that are well-validated but imprecisely targeted to the patient's goals or new measures that are incompletely validated but a good fit with patient priorities. An interesting compromise is a validated questionnaire (MYMOP) that uses standardized wording and response options applied to the symptoms and concerns of greatest interest to the patient.[26]

N-of-1 trials can make use of the entire spectrum of data collection modalities. Traditional approaches include use of surveys, diaries, medical records, and administrative data. Recent developments in information technology have opened the door to several new approaches, including ecological momentary assessment (EMA), and remote positional and physiologic monitoring. Mobile-device EMA cues the patient to input data at more frequent intervals (e.g. hourly, daily, to weekly) than is typical using traditional survey modalities. Compliance with such devices is higher than with paper diaries.[27] When equipped with GPS or actigraphic technology, mobile devices can also track patient movements and activities. Ancillary monitoring devices can be connected to mobile devices to monitor heart rate, blood pressure, blood glucose, Galvanic skin response, electroencephalographic activity, degree of social networking, vocal stress, etc. Data on the reliability and validity of these measures is currently scant but is accumulating rapidly.[28]

Statistical Analysis and Feedback for Decision Making.

Once data are collected, they need to be analyzed and presented to the relevant decision makers in a format that is actionable. In the systematic review by Gabler et al.,[29] approximately half of the trials reported using a t-test or other simple statistical criterion (44%), while 52% reported using a visual/graphical comparison alone. Of the 60 trials (56%) reporting on more than 1 individual, 26 (43%) reported on a pooled analysis. Of these, 23% used Bayesian methodology, while the rest used frequentist approaches to combining the data. Guidance on statistical analytic approaches for n-of-1 trials is left to Chapter 4 (Statistics).

While n-of-1 trials can promote other goals (e.g., increased patient engagement),[9] the primary objective is generally to promote better health care decision making for participating patients. The degree to which decision making can be improved will depend on the quality of the data and the clarity with which results are communicated to the end users, especially the patient

participating in the trial. There are three fundamental issues n-of-1 trialists should consider. First, should outcomes data be presented item-by-item (or scale-by-scale) or as a composite measure? A patient with asthma may be interested in capacity to climb stairs, ability to sleep through the night, and avoidance of the emergency room. These outcomes could be presented as three separate statistics, graphs, or figures. Or they could be combined into a single composite measure that averages the individual components (with or without weighting). The advantage of singlet measures is that they retain clinical granularity and, of themselves, are readily interpretable. The disadvantage is that they can be confusing, especially if different outcomes are affected differently by the treatments under study. The advantage of composite measures is that they make individual-level decision making more straightforward. If, for a given patient, the Asthma Improvement Index moves in a more positive direction on Treatment A than B, the drug of choice is Treatment A. On the other hand, composite outcomes are harder to interpret and may be driven by the most sensitive component (which is not necessarily the most important). In addition, Bayesian analysis of a series of related n-of-1 trials is greatly facilitated by the use of specific, identical outcome measures rather than composites.

The second issue is how to present the data: as graphics, statistics, or both. Simple graphical analysis can transmit results clearly, but not all formats are equally understandable, particularly to low-numeracy populations.[30] In addition, graphical analysis can magnify small differences that a proper statistical analysis would show are likely due to chance. A combined approach may work best, employing statistics to test for stochastic significance (or, using a Bayesian framework, to estimate post-test probabilities) and graphics to lend clarity to the findings.

The third issue is whether to rely solely on the results of the current n-of-1 trial for decision making or to "borrow from strength" by combining current data with the results of previous n-of-1 trials completed by similar patients. The choice will usually be driven by the availability of relevant data and by the ratio of within-patient versus between-patient variance (see Chapter 4

for details). If a similar series of trials has never been conducted, and if few patients have been enrolled in the current series, than by default decision-making rests on the results of the current n-of-1 trial alone. If on the other hand large numbers of patients have completed similar n-of-1 trials, and if within-patient variance is large compared to between-patient variance, then "borrowing from strength" will enhance the precision of the result. Further discussion occurs in Chapter 4 (Statistics).

Opportunities and Challenges

In addition to their potential for enhancing therapeutic precision, n-of-1 trials may offer three broader benefits. First, they may help patients and clinicians recognize ineffective therapies, thus reducing polypharmacy, minimizing adverse effects, and conserving health care resources. Second, they may help engage patients in their own care.[9] A robust literature supports the premise that increased patient involvement in care is associated with better outcomes.[8,31] By helping patients attend to their own outcomes and think critically about treatments, n-of-1 trials can awaken patients' "inner scientist" and give them greater stake in the process of clinical care. Third, n-of-1 trials can blur the boundaries between clinical practice and clinical research, making research more like practice and practice more like research. Making research more like practice is desirable in order to increase the generalizability of clinical research findings. Making practice more like research will create opportunities for developing the clinical evidence base by enhancing systematic data collection on the comparative effectiveness of treatments by real health care professionals treating real patients. As n-of-1 trials become better integrated into practice, a number of other downstream benefits may occur, including:

- Patients become more acquainted with the scientific method and in particular the value of rigorous clinical experiments;

- Clinicians become more connected to the process of generating clinical evidence, more engaged in clinical research, and potentially more interested in participating in clinical trials;
- Practices start collecting data on the relationship between treatments and outcomes, and making such data available for use in routine patient care. If leveraged to full advantage, these data could become the lynchpin of a "learning healthcare system" as envisioned by the Institute of Medicine.[32]

Outline of the Rest of the Monograph

In the rest of this Monograph, authors will expand on themes introduced here. Chapter 2 addresses human subjects issues germane to n-of-1 trials, in particular how n-of-1 trials are situated on the continuum between pure clinical care and pure research. This chapter also provides guidance for IRB chairs and committee members considering applications for running n-of-1 trials. Chapter 3 takes on the very practical issue of how much n-of-1 trials cost, how much value they offer, and what factors organizations should consider before constructing an n-of-1 trial service. Chapter 4 provides an overview of statistical design and analysis considerations, while Chapter 5 outlines key components of information technology infrastructure needed to deploy n-of-1 trials efficiently. Finally, Chapter 6 takes up training and engagement of clinicians and patients preparing to participate in n-of-1 trials.

Checklist

Guidance	Key Considerations	Check
Determine whether n-of-1 methodology is applicable to the clinical question of interest	• Indications include: a) substantial clinical uncertainty; b) chronic or frequently recurring, symptomatic condition; c) treatment with rapid onset and minimal carryover • Contraindications include: a) rapidly progressive condition; b) treatment with slow onset or prolonged carryover; c) patient or clinician insufficiently interested in reducing therapeutic uncertainty to justify effort	☐
Select trial duration, treatment period length, and sequencing scheme	• Longer trial duration delivers greater precision, but completion can be difficult or tedious, with the potential for extended exposure to inferior treatment during trial • Treatment period length should be adjusted to fit the therapeutic half-life (of drug treatments) or treatment onset and duration (of non-drug treatments) • Simple randomization (e.g., AABABBBA) optimizes blinding (more difficult to guess treatment), while balanced sequencing (e.g., ABBABAAB) is a more reliable guarantor of validity	☐
Invoke a suitable washout period, if indicated	• Washout not necessary if treatment duration of action short relative to treatment period • Washout contraindicated if patient could be harmed by cessation of active treatment	☐
Decide whether or not to invoke blinding	• Blinding feasible for some drug treatments but infeasible for most non-drug treatments (behavioral, lifestyle) • Adequate blinding allows investigators to distinguish between specific and nonspecific treatment effects • In some circumstances, this distinction may not matter to patient and clinician; in others, participants may be primarily interested in the combined treatment effect (specific + non-specific)	☐
Select suitable outcomes domains and measures	• Patient preferences pre-eminent, but clinicians' goals and external factors should be accounted for and may occasionally supervene	☐

	• Valid and reliable measures are preferred when available, but patient-centeredness should not be sacrificed to psychometric imperatives.	
Analyze and present data to support clinical decision making by patients and clinicians	• There is a natural tension between identifying a single, primary outcome for decision making and coming to a full understanding of the data • A reasonable approach is to select one or two primary outcome measures but present use a variety of statistical and graphical methods to fully explicate the data.	☐

References

1. Guyatt GH, Haynes RB, Jaeschke RZ, et al. Users' Guides to the Medical Literature: XXV. Evidence-based medicine: principles for applying the Users' Guides to patient care. Evidence-Based Medicine Working Group. *JAMA.* Sep 13 2000;284(10):1290-1296.
2. Perdices M, Tate RL. Single-subject designs as a tool for evidence-based clinical practice: Are they unrecognised and undervalued? *Neuropsychological rehabilitation.* Dec 2009;19(6):904-927.
3. Hogben L, Sim M. The self-controlled and self-recorded clinical trial for low-grade morbidity. *British journal of preventive & social medicine.* Oct 1953;7(4):163-179.
4. Guyatt G SD, Taylor DW, Chong J, Roberts R, Pugsley S. Determining optimal therapy--randomized trials in individual patients. *N Engl J Med.* 1986;314(14):889-892.
5. Guyatt G SD, Adachi J, Roberts R, Chong J, Rosenbloom D, Keller J. A clinician's guide for conducting randomized trials in individual patients. *Canadian Medical Association Journal (CMAJ)* 1988 Sep 15;139(6):497-503.
6. Guyatt GH HA, Jaeschke R, Keller J, Adachi JD, Roberts RS. N of 1 randomized trials for investigating new drugs. *Control Clin Trials. .* 1990 Apr;;11(2):88-100.
7. Kravitz RL DN, Niedzinski EJ, Hay MC, Subramanian SK, Weisner TS. What ever happened to N-of-1 trials? Insiders' perspectives and a look to the future. *Milbank Q.* 2008 Dec;;86(4):533-555.
8. Greenfield S, Kaplan S, Ware JE, Jr. Expanding patient involvement in care. Effects on patient outcomes. *Annals of internal medicine.* Apr 1985;102(4):520-528.
9. Nikles CJ CA, Del Mar CB. Using n-of-1 trials as a clinical tool to improve prescribing. *Br J Gen Pract. .* 2005 Mar;;55(512):175-180.
10. Greenfield S KR, Duan N, Kaplan SH. Heterogeneity of treatment effects: implications for guidelines, payment, and quality assessment. *Am J Med.* 2007 Apr;;120(4 Suppl 1):S3-9.
11. Kravitz RL DN, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Q.* 2004;82(4):661-687.
12. Kent DM, Rothwell PM, Ioannidis JP, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials.* 2010;11:85.
13. Rothwell PM, Mehta Z, Howard SC, Gutnikov SA, Warlow CP. Treating individuals 3: from subgroups to individuals: general principles and the example of carotid endarterectomy. *Lancet.* Jan 15-21 2005;365(9455):256-265.
14. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet.* Jan 1-7 2005;365(9453):82-93.
15. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet.* Jan 8-14 2005;365(9454):176-186.
16. Warden D, Rush AJ, Trivedi MH, Fava M, Wisniewski SR. The STAR*D Project results: a comprehensive review of findings. *Current psychiatry reports.* Dec 2007;9(6):449-459.
17. De Gruttola VG, Clax P, DeMets DL, et al. Considerations in the evaluation of surrogate endpoints in clinical trials. summary of a National Institutes of Health workshop. *Controlled clinical trials.* Oct 2001;22(5):485-502.
18. Kravitz RL DN, White RH. N-of-1 Trials of Expensive Biological Therapies: A ThirdWay? . *Archives of Internal Medicine* 2008;168:1030–1033.
19. S. S. Cross-over trials in clinical research. . 1993.
20. S. S. Individual response to treatment: is it a valid assumption? *BMJ.* 2004 Oct 23;;329(7472):966-968.
21. Matthews JN. Multi-period crossover trials. *Statistical methods in medical research.* Dec 1994;3(4):383-405.

22. Ulmer M, Robinaugh D, Friedberg JP, Lipsitz SR, Natarajan S. Usefulness of a run-in period to reduce drop-outs in a randomized controlled trial of a behavioral intervention. *Contemporary clinical trials.* Sep 2008;29(5):705-710.

23. Benedetti F, Mayberg HS, Wager TD, Stohler CS, Zubieta JK. Neurobiological mechanisms of the placebo effect. *The Journal of neuroscience : the official journal of the Society for Neuroscience.* Nov 9 2005;25(45):10390-10402.

24. Agarwal R, Bills JE, Hecht TJ, Light RP. Role of home blood pressure monitoring in overcoming therapeutic inertia and improving hypertension control: a systematic review and meta-analysis. *Hypertension.* Jan 2011;57(1):29-38.

25. Lin EH, Von Korff M, Ciechanowski P, et al. Treatment adjustment and medication adherence for complex patients with diabetes, heart disease, and depression: a randomized controlled trial. *Annals of family medicine.* Jan-Feb 2012;10(1):6-14.

26. Paterson C. Measuring outcomes in primary care: a patient generated measure, MYMOP, compared with the SF-36 health survey. *BMJ.* Apr 20 1996;312(7037):1016-1020.

27. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. *Controlled clinical trials.* 04/ 2003;24(2):182-199.

28. Ramanathan N, Alquaddoomi F, Falaki H, et al. ohmage: An open mobile system for activity and experience sampling. Paper presented at: Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2012 6th International Conference on; 21-24 May 2012, 2012.

29. Gabler NB, Duan N, Vohra S, Kravitz RL. N-of-1 trials in the medical literature: a systematic review. *Medical care.* Aug 2011;49(8):761-768.

30. Ancker JS, Senathirajah Y, Kukafka R, Starren JB. Design Features of Graphs in Health Risk Communication: A Systematic Review. *Journal of the American Medical Informatics Association.* November 1, 2006 2006;13(6):608-618.

31. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. *JAMA.* Nov 20 2002;288(19):2469-2475.

32. Medicine) IIo. *Best care at lower cost: the path to continuously learning healthcare in America.* Washington, DC: National Academy Press; 2012.

Figure 1. Design of a prototypical n-of-1 trial.